Characterisation of dairy strains of *Geobacillus stearothermophilus* and a genomics insight into its growth and survival during dairy manufacture

A thesis presented in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy
in Microbiology
at Massey University, Palmerston North,
New Zealand

Sara Burgess

2016

This thesis is dedicated to my sons, Samuel and James.

Abstract

The thermophilic bacilli, such as *G. stearothermophilus*, are an important group of contaminants in the dairy industry. Although these bacilli are generally not pathogenic, their presence in dairy products is an indicator of poor hygiene and high numbers are unacceptable to customers. In addition, their growth may result in milk product defects caused by the production of acids or enzymes, potentially leding to off-flavours. These bacteria are able to grow in sections of dairy manufacturing plants where temperatures reach 40 – 65 °C. Furthermore, because they are spore formers, they are difficult to eliminate. In addition, they exhibit a fast growth rate and tend to readily form biofilms. Many strategies have been tested to prevent the formation of thermophilic bacilli biofilms in dairy manufacture, but with limited success. This is, in part, because little is known about the diversity of strains found in dairy manufacture, the structure of thermophilic bacilli biofilms and how these bacteria have adapted to grow in a dairy environment.

In Chapters 2 and 3, phenotypic approaches were taken to understand the diversity of strains within a manufacturing plant. Specifically in Chapter 2, strains of the most dominant thermphilic bacilli, *G. stearothermophilus*, were isolated from the surface of various locations within the evaporator section and ten strains were evaluated for different phenotypic characteristics. Biochemical profiling, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and fatty profiling demonstrated that the population was diverse. In Chapter 3, it was shown that the same ten strains varied in their ability to form biofilms and produce spores. Three strains of *G. stearothermophilus*, A1, P3 and D1, were selected for further analysis. SEM demonstrated that there were differences in biofilm morphologies between the three strains, particularly D1 versus the other two strains, A1 and P3.

In Chapters 4, 5 and 6 a comparative genomics approach was taken to determine how these bacteria are able to grow and survive within a dairy manufacturing environment, as well as how they differ from other strains of *Geobacillus*. In Chapter 4 draft genome sequences were generated for three strains of *G.stearothermophilus.* Identification of a putative lactose operon in the three dairy strains provided evidence of dairy adaptation. In Chapter 5 a phylogenomics approach was taken to resolve relationships within the *Geobacillus* genus and to identify differences within the *G. stearothermophilus* group itself. Finally in Chapter 6 comparison with the model organism *B. subtilis,* gave a genomics insight into the potential mechanisms of sporulation for *Geobacillus* spp.

List of Publications

Burgess S A, Flint S H, Lindsay D, Cox M P and Biggs P J (2016). An updated analysis of *Geobacillus* taxonomy based on phylogenomic principles. Submitted to *BMC Microbiology*.

Burgess S A, Cox M P, Flint S H, Lindsay D and Biggs P J (2015). Draft genome sequences of three strains of *Geobacillus stearothermophilus* from a milk powder manufacturing plant. *Genome Announcements*. doi:10.1128/genomeA.00939-15.

Burgesss S A, Lindsay D and Flint S H (2014). Biofilms of thermophilic bacilli isolated from dairy processing plants and efficacy of sanitisers. In: *Microbial Biofilms: Methods and Protocols*, Methods in Molecular Biology, vol. 1147, pp 367-377, Springer, New York.

Burgess S A, Flint S H and Lindsay D (2014). Characterization of thermophilic bacilli from a milk powder processing plant. *Journal of Applied Microbiology*. 11: 350-359.

Burgess S A, Lindsay D, Flint S H (2010). Thermophilic bacilli and their importance in dairy processing. *International Journal of Food Microbiology*. 144: 215 – 225.

List of Presentations

Burgess S A, Lindsay D, Flint S H and Biggs P J (2015). Genomic insights into the adaptation of *Geobacillus stearothermophilus* to a milk powder manufacturing plant. Presented at: *New Zealand Microbiological Society Conference, Rotorua, November 2016*.

Burgess S A, Lindsay D, and Flint S H (2012). Biofilm formation and spore production of *Geobacillus stearothermophilus*. Presented at: *New Zealand Microbiological Society Conference, Dunedin, November 2012*.

Acknowledgements

To my supervisors Steve Flint, Denise Lindsay, and Roberto Kolter thank-you for sticking with me over the 6 years it took to complete this journey. Steve, you always have an amazing amount of enthusiasm. Denise, thank-you for your thorough reviews and providing such a supportive environment when things got tough. Roberto, I will be forever grateful for taking me on even though you had never heard of me. Your probing questions at meetings kept me on my toes and the time I spent in your lab opened my eyes to a different world of microbiology.

To my additional supervisor Patrick Biggs, who came on board part-way through my PhD, thank-you for taking me on. Your expertise was a lifesaver and your enthusiasm gave me the boost I needed. I have learnt an enormous amount from you.

Steve Holroyd, without your support I never would have started this PhD. Thank-you for being there as a sounding-board when things got tough.

Hera Vlamakis, you are one of the most wonderful people I've met. Thank-you for making me so welcome in the Kolter lab and for not rolling your eyes when going through some basic microbial genetics.

To everyone in the Kolter lab (Erin, Pascale, Beth, Matt, Diana, Robe, Majse, Christian, Liraz, Maerite, Marta, Diego, and Ilana) thank-you for making me so welcome and providing advice when needed.

Thank-you to Fulbright New Zealand for giving me the opportunity to go to Boston through the Fulbright Science and Innovation Graduate Award.

Ashlee Earle, thank-you for getting me started with the comparative genomics.

Murray Cox, thank-you for generating the draft assemblies. I also really appreciated the thorough reviews of some of my draft manuscripts with some probing questions that got me thinking more outside the square.

Teng Yi Sin, the Massey Singapore student, thank-you for your dedication and hard-work during your project.

Doug and Jordan from the Massey University microscopy unit thank-you for your help and advice with the microscopy imaging.

Jasna Rakonjac and Wesley Wen thank-you for your advice and help with DNA extractions. Jasna I also really appreciated being able to come to your laboratory meetings and for giving advice when needed.

Anne-Marie Jackson and Julia Good I really appreciated you making me feel at ease when moving to the SEAT microbiology lab and helping me out with equipment.

Matt and Steve thank-you for the IT support.

Thank-you to the various people at the Fonterra Cooperative Group who helped out with laboratory work and sampling, including Donna Crowe, Matt Lin, Jacinda Aplin, Saloni Gokarn, Mark Reynolds, Helaina Colle, Micro-servicing staff and plant operators from Kauri and Te Rapa.

To my family I could not have done this without your love and support. To my Dad for initiating some ideas, being there when I needed a sounding board or advice. To my Mum for looking after the kids sometimes, for the dinners every Tuesday and hugs when I needed them.

To my husband Mitch, thank-you so so so much. You have been so supportive, you gave up your job so that I could spend time in the Kolter lab, you have been there for the kids at night so that I could get my sleep and you even built me a computer that made the genomic analyses so much easier. I could not have done this without you. Finally to my two sons Samuel and James, who tolerated many mornings without me.

Significant contributions from others

- Fatty acid profiling by Saloni Gokarn and Mark Reynolds, Fonterra Research and Development Centre
- 18-h biofilm assays by Massey University Singapore student Teng Yi Sin
- Biofilm SEM imaging by Massey University microscopy unit (MMIC)
- Mi-Seq™ sequencing by New Zealand Genomics Ltd., Massey Genome Service, Massey University
- Draft genome assemblies by Murray Cox
- In-house Perl scripts and reputer analysis showing repetative regions in the genome of *G. kaustophilus* HTA426 prepared by Patrick Biggs

Table of Contents

List of Figures

List of Tables

Non-standard abbreviations

A_w	Water activity
ANI	Average nucleotide identity
BDBH	Bidirectional best hit
CIP	Clean-in-place
COG	Clusters of orthologous groups
DPA	Dipicolinic acid
DSI	direct steam injection
EOR	End-of-run
GFF	General file format
HK	Histidine kinase
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
MCL	Markov clustering
MLVA-HRM	Multi-locus variable-number analysis - high-resolution melt analysis
MVR	Mechanical vapour recompression (evaporator)
n/d	Not determined
PHE	Plate heat exchanger
rMLST	Ribosomal multi-locus sequence typing
SFB	Static fluid bed
SOR	Start-of-run
THE	Tubular heat exchanger
TM	Transmembrane

T/S	Total solids
TVR	Thermal vapour recompression (evaporator)

Definitions

Accessory genome	Additional genes that are present in some members and absent from others within a group of isolates under investigation.
Clean-in-place (CIP)	Cleaning regime after a manufacturing run.
Conditioning layer	The thin layer of proteins and exopolysaccharides that forms immediately on a surface when it is submersed in a liquid.
Core genome	A set of genes shared by all members in a group of isolates under investigation.
Direct steam injection (DSI)	A direct method of heat treatment where steam is injected into the milk.
Effect	A section of an evaporator that has the same boiling temperature.
Engulfment	Part of the sporulation process where degradation of the septal membrane (between the mother-cell and forespore), and relocation of the mother-cell membrane around the forespore occurs.
Forespore	The immature form of the spore when it is being formed within the mother cell.
Foulant	The build-up of milk proteins and calcium phosphate salts on equipment surfaces in dairy manufacturing plants.
Homologue	Genes that are descendents of the same ancestral gene but were separated by either speciation or gene duplication.
Orifice pans	Located at the top of the evaporator to distribute milk into the pass tubes.
Mother cell	The cell which houses the forespore as it matures into an endospore.
Orthologue	Genes in different species which were derived from the same ancestral gene and were separated by speciation.
Paralogue	Genes that are descendents of the same ancestral gene, but were separated by gene duplication.
Pass	A section of the effect, in the evaporator, that is made up of a set of

	tubes that the milk passes through.
Plate heat exchanger (PHE)	An indirect method of heat treatment that consists of a series of plates where the heating or cooling medium passes on one side, and the milk on the other.
Pseudogene	A “gene” which has lost its ability to code for a protein, generally through the accrual of mutations such as stop codons or frameshifts within the gene.
Sliding	The passive movement of bacteria across a surface. This process does not make use of bacterial appendages such as flagella or pilli.
Spore coat	The outer layers of the endospore.
Spore cortex	The layer between the inner and outer membranes of the spore and is composed of peptidoglycan.
Spore crust	The outer layer of the coat in spores of *Bacillus subtilis.*
Spore exosporium	The outermost layer of spores in some species of *Bacillus*. It is composed of glycoprotein and separated from the coat by a large irregular space.
Swarming	The coordinated movement, through the use of flagella, of a bacterial population across a surface.
Water activity	In the dairy context this refers to the amount of water not bound to food molecules. This water can enable the growth of bacteria. When milk powder is made the water activity decreases through the evaporators as the milk is concentrated and once dried reaches levels that no longer supports bacterial growth.

Description of computer programs and on-line genomic tools

Bowtie 2	An alignment program used for aligning short sequences (e.g. sequence reads from a genome sequencer) to long sequences (e.g. genome sequences) (Langmead & Salzberg, 2012). The output generated by Bowtie 2 is a SAM file.
COGnitor	A software tool designed to assign predicted proteins to the already established COGs (Tatusov *et al.*, 2000, Galperin *et al.*, 2015).
CRISPRDetect	An on-line tool (http://brownlabtools.otago.ac.nz/CRISPRDetect/predict_crispr_array.html), designed to detect the presence of CRISPR arrays (Biswas *et al.*, 2014).
CRISPRTarget	An on-line tool (http://brownlabtools.otago.ac.nz/CRISPR_WEB/crispr_analysis.html), designed to determine the target of CRISPR spacers (Biswas *et al.*, 2013).
GET_HOMOLOGUES	A software package that incorporates three different algorithms (BDBH, COGtriangles and OrthoMCL) for clustering homologous genes (Contreras-Moreira & Vinuesa, 2013).
Jspecies	A software package designed for comparing the similarity of two or more bacterial species (Richter & Rossello-Mora, 2009). Synthetic DNA-DNA hybridisations can be carried out using three methods: Average nucleotide identity (ANI) calculated using BLAST, ANI calculated using MUMmer and calculation of tetra nucleotide frequencies (TETRA).
OrthoMCL	A software program which uses an algorithm incorporating both BLASTP and the Markov clustering algorithm to determine orthologous groups of proteins within a group of genomes (Li *et al.*, 2003).
Pfam	Pfam (http://pfam.xfam.org/) is a database of protein families. In this present study it was used for identifying domains in predicted protein sequences.
Prokka	Prokka is a software package used for rapidly annotating prokaryotic

	genomes (Seemann, 2014).
Rapid Annotation using Subsystem Technology (RAST)	A web based server (http://rast.nmpdr.org/), which can carry out automated annotations on bacterial genomes (Aziz *et al.*, 2008).
RNAmmer	A web based server (http://www.cbs.dtu.dk/services/RNAmmer/), used to predict prokaryotic and eukaryotic rRNA gene sequences in genome sequences (Lagesen *et al.*, 2007).
Velvet	An algorithm package used for *de novo* genome assembly (Zerbino & Birney, 2008). In assembling, the sequence reads are broken into shorter sequences called k-mers and used to generate de Bruijn graphs. A range of k-mer lengths are tested to generate the best assembly.